

Estimation of Hurst Parameter of Fractional Brownian Motion Using CMARS Method

F.Yerlikaya-Özkurt^{a,*}, C.Vardar-Acar^b, Y.Yolcu-Okur^a, G.-W.Weber^a

^a*Institute of Applied Mathematics, METU, 06800 Ankara, Turkey*

^b*Department of Mathematics, TOBB ETU, 06530 Söğütözü, Ankara, Turkey*

Abstract

In this study, we develop a new theory of estimating Hurst parameter using conic multivariate adaptive regression splines (CMARS) method. We concentrate on the strong solution of stochastic differential equations (SDEs) driven by fractional Brownian motion (fBm). The superiority of our approach to the others is, it not only estimates the Hurst parameter but also finds spline parameters of the stochastic process in an adaptive way. We examine the performance of our estimations using simulated test data.

Keywords: Stochastic differential equations, fractional Brownian motion, Hurst parameter, conic multivariate adaptive regression splines

2000 MSC: 60G22, 60H10, 90C20, 90C90

1. Introduction

Fractional Brownian motion (fBm) is a widely used concept for modelling various situations such as the level of water in a river, the temperature at a specific place, empirical volatility of a stock, the price dynamics of electricity. It appears naturally in these phenomena because of its capability of explaining the *dependence* structure in real-life observations. This structure in fBm is represented by its Hurst parameter H . A fBm with Hurst parameter $H > 1/2$ is called a *persistent* process, i.e., the increments of this

*Corresponding author

Email addresses: fatmayerlikaya@gmail.com (F.Yerlikaya-Özkurt),
cvardar@etu.edu.tr (C.Vardar-Acar), yyolcu@metu.edu.tr (Y.Yolcu-Okur),
gweber@metu.edu.tr (G.-W.Weber)

process are positively correlated. On the other hand, the increments of a fBm with $H < 1/2$ is called an *anti-persistent* process with increments being negatively correlated. For $H = 1/2$, fBm corresponds to Brownian motion which has independent increments. For further information on fBm and its applications, see [2, 13, 17].

It is highly important to identify the value of Hurst parameter in order to understand the structure of the process and its applications since the calculations dramatically differ according to the value of H . Therefore, some techniques have been developed to estimate Hurst parameter which can be categorized into three groups; heuristics, maximum likelihood and wavelet-based estimators. In the group of heuristics estimators, there is R/S estimator which was firstly proposed by Hurst [8], followed by the methods of correlogram, variogram, variance plot, and partial correlations plot. Due to lack of accuracy of heuristics estimators, maximum likelihood estimators (mle) were developed. Being weakly consistent is the main disadvantage of mle. In paralel to mle, wavelet-based estimators were suggested because of the popularity of wavelet decomposition of fBm [3, 18].

In search of faster and efficient ways to estimate the Hurst parameter H , we suggest a new numerical and computational way, *conic multivariate adaptive regression splines (CMARS)*. CMARS is an alternative approach to the well-known data mining tool *multivariate adaptive regression splines (MARS)*. It is based on a penalized residual sum of squares (PRSS) for MARS as a Tikhonov regularization (TR) problem. CMARS treats this problem by a continuous optimization technique, in particular, the framework of *conic quadratic programming (CQP)*. These convex optimization problems are very well-structured, herewith resembling linear programs and, hence, permitting the use of *interior point methods*.

This paper is organized as follows; in Section 2, we start with explaining the properties of our madel given as, SDEs driven by fBm. In Section 3, we introduce the method CMARS relating it to the Hurst parameter estimation of our model. In Section 4, we give an application of our study, in order to test the theory we have developed. Finally, we present a brief conclusion and a general outlook of our study.

2. Stochastic Differential Equations with Fractional Brownian Motion

Stochastic Differential Equations (SDEs) generated by fBm are widely used to represent noisy and real-world problems. They play an important role in many fields of science such as finance, physics, biotechnology and engineering. In this section, we briefly recall some concepts on fBm and stochastic differential equations driven by fBm.

2.1. Fractional Brownian Motion

Let H be a constant in the interval $(0, 1)$. FBm $(W^H(t))_{t \geq 0}$ with Hurst parameter H , is a continuous and centered Gaussian process with covariance function

$$E[W^H(t)W^H(s)] = \frac{1}{2}(t^{2H} + s^{2H} - |t - s|^{2H}).$$

We note that, for $H = 1/2$, fBm corresponds to a standard Brownian motion which has independent increments. For a *standard fBm*, $W^H(t)$:

- $W^H(0) = 0$ and $E[W^H(t)] = 0$ for all $t \geq 0$.
- W^H has homogenous increments, i.e., $W^H(t+s) - W^H(s)$ has the same law as $W^H(t)$, for all $s, t \geq 0$.
- W^H is a Gaussian process and $E[(W^H(t))^2] = t^{2H}$ ($t \geq 0$), for all $H \in (0, 1)$.
- W^H has continuous trajectories.

The Hurst parameter H of fBm explains the dependency of data. Indeed, the correlation between increments for $s, t \geq 0$ can be obtained by;

$$E[(W^H(t+h) - W^H(t))(W^H(s+h) - W^H(s))] = \frac{h^{2H}}{2}[(n+1)^{2H} + (n-1)^{2H} - 2n^{2H}].$$

It can be seen that observations with $H > 1/2$ have positively correlated increments and display long-range dependence, while the observations with $H < 1/2$ have a negatively correlated increments and display a short-range dependence structure (see Figure 1). Therefore, it is crucial to find the Hurst parameter of a stochastic process for understanding the structural behaviour of this phenomena. In this study, we concentrate on finding H for the stochastic processes which are the strong solutions of SDEs with fBm. Hence, we first recall some fundamental properties of them.

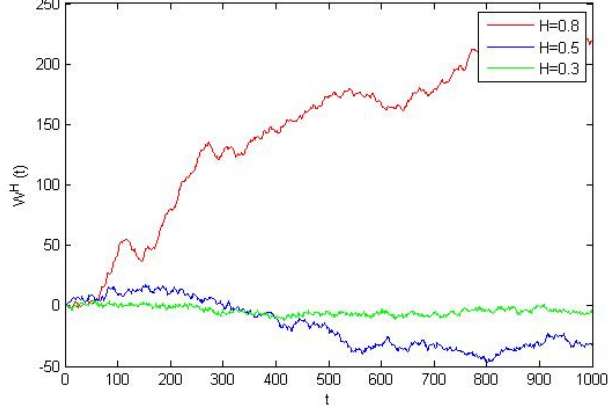


Figure 1: Sample paths of fBm with different Hurst parameter values simulated by Cholesky method.

2.2. Stochastic Differential Equations Driven by Fractional Brownian Motion

Suppose we have a stochastic process $X = \{X(t); t \geq 0\}$ defined on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$ which is the strong solution of the following SDE:

$$dX(t) = a(t, X(t))dt + b(t, X(t))dW^H(t). \quad (1)$$

Here, $a(t, X(t))$ and $b(t, X(t))$ are the drift and diffusion terms satisfying the conditions of existence and uniqueness theorem for $t \geq 0$. Note that it is necessary to have the integrator as a semi-martingale in the theory of stochastic integration. However, since fBm is not a semi-martingale, one should extend the usual settings as in Itô integral and define the integration with respect to fBm in a new pathwise integration technique. Alos et al. [1] construct the theory of integration with respect to general Gaussian processes to overcome this. For further studies on this topic, see [13, 17].

There have been comprehensive studies on statistical inferences for processes satisfying SDEs driven by Brownian motion. However, the recent interest is on SDEs driven by fBm since there have not been adequate studies on this topic. The purpose of this study is to estimate the Hurst parameter of the following SDE

$$dX(t) = a(t, X(t))dt + b dW^H(t), \quad X_0 \in \mathbb{R}, \quad (2)$$

by *Conic Multivariate Adaptive Regression Splines (CMARS)* methodology. Note that $b(t, X(t)) \equiv b$ term in equation (1) is taken as *constant*.

3. Estimation of Hurst Parameter Using Conic Multivariate Adaptive Regression Splines Method

In this section, as an alternative to the existing methods of estimation of Hurst parameter, CMARS and the related proposed methodology will be introduced. For that purpose, firstly, we give a brief description of CMARS method and then we mention about the methodology and show how to apply this technique for finding the Hurst parameter of SDE defined in equation (2).

3.1. Method of Conic Multivariate Adaptive Regression Splines

CMARS method is an alternative approach to the well-known data mining tool *Multivariate Adaptive Regression Splines (MARS)*. It makes no specific assumption about the underlying functional relationship between the dependent and independent variables to estimate a general model function [5]. CMARS is introduced by linear combinations of the basis functions (BFs) that are used in MARS. The selection of BFs is data-based and specific to the problem at hand. CMARS uses one-dimensional BFs of the form $c^+(x, \tau) = [(x - \tau)]_+$ and $c^-(x, \tau) = [-(x - \tau)]_+$, where $[q]_+ := \max\{0, q\}$ (see [20, 22] for further details). Each function is piecewise linear, with a knot at the value τ , and the corresponding couple of function is called a *reflected pair*. A set of BFs is given as follows:

$$\wp := \{(x_j - \tau)_+, (\tau - x_j)_+ \mid \tau \in \{x_{1,j}, x_{2,j}, \dots, x_{N,j}\}, j \in \{1, 2, \dots, p\}\}.$$

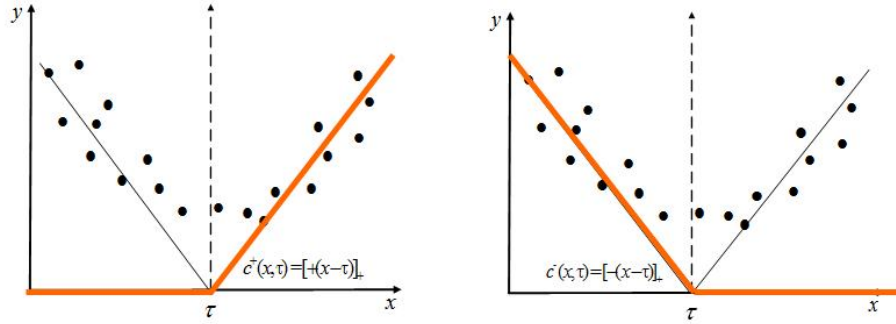


Figure 2: Visualization of the BFs.

A CMARS model function f is represented by a linear combination of BFs which is successively built up by the set \wp as described below:

$$Y = f(\mathbf{x}) + \epsilon = \theta_0 + \sum_{m=1}^M \theta_m \psi_m(\mathbf{x}^m) + \epsilon. \quad (3)$$

Here Y is a response variable, $\mathbf{x}^m = (x_1, x_2, \dots, x_p)^T$ a vector of predictors for the corresponding m th multivariate basis function. Furthermore, θ_m are the unknown coefficients for the m th basis function ($m = 1, 2, \dots, M$) or for the constant 1 ($m = 0$), and ϵ is an additive stochastic component which is assumed to have zero mean and finite variance. In equation (3), ψ_m ($m = 1, 2, \dots, M$) are BFs as products of two or more one-dimensional BFs. Such interaction BFs are created by multiplying an existing basis function with a truncated linear function, involving a new variable. The form of the m th BF can be written as follows:

$$\psi_m(\mathbf{x}^m) := \prod_{j=1}^{K_m} [s_{\kappa_j^m} \cdot (x_{\kappa_j^m} - \tau_{\kappa_j^m})]_+. \quad (4)$$

Here, \mathbf{x}^m is the vector of variable contributed to the m th BF, K_m is the number of truncated linear functions multiplied in the m th BF, $x_{\kappa_j^m}$ is the predictor variable corresponding to the j th truncated linear function in the m th BF, $\tau_{\kappa_j^m}$ is the knot value corresponding to the variable $x_{\kappa_j^m}$, and $s_{\kappa_j^m}$ is the selected sign +1 or -1.

CMARS is constructed by a *Penalized Residual Sum of Squares (PRSS)* parameter estimation problem, instead of an ordinary least-squares estimation problem as it occurs in MARS method. The PRSS problem aims at accuracy and a smallest possible complexity of the model. PRSS with penalty parameters λ_m and with M_{max} BFs which are accumulated in the first part of the MARS algorithm, has the following form:

$$PRSS := \sum_{i=1}^N (y_i - f(\tilde{\mathbf{x}}_i))^2 + \sum_{m=1}^{M_{max}} \lambda_m \sum_{\substack{|\alpha|=1 \\ \alpha=(\alpha_1, \alpha_2)^T}}^2 \sum_{\substack{r \leq s \\ r, s \in V_m}} \int_{Q^m} \theta_m^2 [D_{r,s}^\alpha \psi_m(\mathbf{t}^m)]^2 d\mathbf{t}^m, \quad (5)$$

where $V_m := \{\kappa_j^m | j = 1, 2, \dots, K_m\}$ is the variable set associated with the m th BF, $\mathbf{t}^m = (t_{m_1}, t_{m_2}, \dots, t_{m_{K_m}})^T$ represents the vector of variables which

contribute to the m th BF. Moreover, $D_{r,s}^{\alpha} \psi_m(\mathbf{t}^m) := \frac{\partial^{|\alpha|} \psi_m}{\partial^{\alpha_1} t_r^m \partial^{\alpha_2} t_s^m}(\mathbf{t}^m)$ for $\alpha = (\alpha_1, \alpha_2)^T$, $|\alpha| := \alpha_1 + \alpha_2$, where $\alpha_1, \alpha_2 \in \{0, 1\}$.

After using the same penalty parameter $\lambda = \lambda_m$ for each derivative, PRSS turns into a *Tikhonov regularization problem* as described below:

$$PRSS \approx \left\| \mathbf{y} - \boldsymbol{\psi}(\tilde{\mathbf{d}})\boldsymbol{\theta} \right\|_2^2 + \lambda \|\mathbf{L}\boldsymbol{\theta}\|_2^2, \quad (6)$$

where \mathbf{L} is constructed by the discretizations of the high-dimensional integrals given in equation (5). The model approximations as presented in equation (6) are carefully prepared. They play an important role in order to raise a final model approximation which is linear in the unknown spline parameters. After unifying some discretised complexity terms and including them into inequality constraints, a *Conic Quadratic Programming (CQP)* problem is obtained which uses *interior point methods* [15, 16]. The formulation of CQP is given as follows:

$$\begin{aligned} \min_{t, \boldsymbol{\theta}} \quad & t, \\ \text{subject to} \quad & \left\| \mathbf{y} - \boldsymbol{\psi}(\tilde{\mathbf{d}})\boldsymbol{\theta} \right\|_2 \leq t, \quad \|\mathbf{L}\boldsymbol{\theta}\|_2 \leq \sqrt{\widetilde{M}}, \end{aligned} \quad (7)$$

referring to some chosen complexity bound $\widetilde{M} \geq 0$.

3.2. Discretization of Stochastic Differential Equations with Fractional Brownian Motion

In general, the distribution of the stochastic process $\{X(t); t \geq 0\}$ is not known. Therefore, the discretized version of the SDEs, $(\hat{X}_i)_{i \in \mathbb{N}}$, should be simulated [10]. There are many discretization schemes for the SDEs generated by fBm such as Euler and Milstein Scheme (see [6] for further details). In this study, *Euler approximation* is used since Milstein approximation contains the derivatives of the diffusion term in equation (2) which is equal to zero. The Euler approximation of the equation (2) is:

$$\hat{X}_{i+1} = \hat{X}_i + a(\hat{X}_i, t_i)(t_{i+1} - t_i) + b(\hat{X}_i, t_i)(W_{i+1}^H - W_i^H) \quad (i \in \mathbb{N}). \quad (8)$$

For finitely many given data points (\bar{X}_i, \bar{t}_i) ($i = 1, 2, \dots, N$), the symbolic form of the *approximation* can be given as follows:

$$\dot{\bar{X}}_i = a(\bar{X}_i, \bar{t}_i) + b(\bar{X}_i, \bar{t}_i) \frac{\Delta W_i^H}{\bar{h}_i}, \quad (9)$$

where $\Delta W_i^H = W_{i+1}^H - W_i^H$ is a centered Gaussian random variable, $\bar{h}_i = \bar{t}_{i+1} - \bar{t}_i := \Delta \bar{t}_i$ represents step lengths and

$$\dot{\bar{X}}_i = \begin{cases} \frac{\bar{X}_{i+1} - \bar{X}_i}{\bar{h}_i} & \text{if } i = 1, 2, \dots, N-1, \\ \frac{\bar{X}_N - \bar{X}_{N-1}}{\bar{h}_N} & \text{if } i = N, \end{cases}$$

represents the difference quotients raised on the i th data value. A more compact form of the equation (9) is defined by

$$\dot{\bar{X}}_i = \bar{G}_i + \bar{F}_i c_i, \quad (10)$$

where $\bar{G}_i := a(\bar{X}_i, \bar{t}_i)$, $\bar{F}_i := b(\bar{X}_i, \bar{t}_i)$, and $c_i := \Delta W_i^H / \bar{h}_i$. Note that equation (10) can be considered as an approximation of the problem. The expressions stated until the end of Section 3 are described parametrically with respect to the Hurst parameter H . In Section 4, we shall specify it by numeric values.

3.3. Parameter Estimation

To determine the unknown values in equation (10), the following minimization problem is constructed using some abbreviated notation of the approximation [19]:

$$\min_{\boldsymbol{\theta}} \sum_{i=1}^N \left\| \dot{\bar{X}}_i - (\bar{G}_i + \bar{F}_i c_i) \right\|_2^2.$$

Here, $\boldsymbol{\theta}$ comprises all unknown parameters in the Euler approximation. To solve this optimization problem and to give a smoother, regularized approximation to the data, we employ CMARS method which controls any high “variation” in the data. CMARS’ BF’s are gradually constructed for the approximation of \bar{G}_i and \bar{F}_i with data $\bar{\mathbf{U}}_{i,B}^l$, $\bar{\mathbf{U}}_{i,C}^m = (\bar{X}_i, \bar{t}_i)$ according to the following approaches [21]:

$$\bar{G}_i = \alpha_0 + \sum_{l=1}^{d^B} \alpha_l B_l(\bar{\mathbf{U}}_{i,B}^l), \quad \text{and} \quad \bar{F}_i c_i = \beta_0 + \sum_{m=1}^{d^C} \beta_m C_m(\bar{\mathbf{U}}_{i,C}^m).$$

Here, the forms of the BF's are $B_l(\bar{\mathbf{U}}_B^l) = \prod_{k=1}^2 [s_{\kappa_k^l}^B \cdot (x_{\kappa_k^l}^B - \tau_{\kappa_k^l}^B)]_+$ and $C_m(\bar{\mathbf{U}}_C^m) = \prod_{k=1}^2 [s_{\kappa_k^m}^C \cdot (x_{\kappa_k^m}^C - \tau_{\kappa_k^m}^C)]_+$ as we described in equation (4). Here, we choose the numbers K_l^B and K_m^C (in sense of equation (4)) as maximal, namely, as 2.

We construct the penalized residual sum of squares (PRSS) for our minimization problem in the following form:

$$\begin{aligned}
PRSS := & \sum_{i=1}^N (\dot{X}_i - (\bar{G}_i + \bar{F}_i c_i))^2 + \\
& \sum_{l=1}^{d^B} \lambda_l \sum_{\substack{|\alpha|=1 \\ \alpha=(\alpha_1, \alpha_2)^T}}^2 \sum_{\substack{r < s \\ r, s \in V_l}} \int_{Q_l^B} \alpha_l^2 [D_{r,s}^\alpha B_l(\bar{\mathbf{U}}_B^l)]^2 d\mathbf{U}_B^l + \\
& \sum_{m=1}^{d^C} \mu_m \sum_{\substack{|\alpha|=1 \\ \alpha=(\alpha_1, \alpha_2)^T}}^2 \sum_{\substack{r < s \\ r, s \in V_m}} \int_{Q_m^C} \beta_m^2 [D_{r,s}^\alpha C_m(\bar{\mathbf{U}}_C^m)]^2 d\mathbf{U}_C^m. \quad (11)
\end{aligned}$$

Here, the multipliers $\lambda_l, \mu_m \geq 0$ are smoothing parameters and they provide a tradeoff between both accuracy and complexity. To approximate two multi-dimensional integrals in equation (11), parallepipies $Q_l^B = [a_{1,B}^l, b_{1,B}^l] \times [a_{2,B}^l, b_{2,B}^l] = \prod_{k=1}^2 Q_{k,B}^l$ and $Q_m^C = [a_{1,C}^m, b_{1,C}^m] \times [a_{2,C}^m, b_{2,C}^m] = \prod_{k=1}^2 Q_{k,C}^m$ which encompass all our input data are constructed. Then, the following discretization is applied for the first multi-dimensional integral:

$$\begin{aligned}
\int_{Q_l^B} \alpha_l^2 [D_{r,s}^\alpha B_l(\bar{\mathbf{U}}_B^l)]^2 d\mathbf{U}_B^l & \approx \sum_{i=1}^{(N+1)^2} \left(\sum_{\substack{|\alpha|=1 \\ \alpha=(\alpha_1, \alpha_2)^T}}^2 \sum_{\substack{r < s \\ r, s \in V_m}} \alpha_l^2 [D_{r,s}^\alpha B_l(\hat{\mathbf{U}}_{i,B}^l)]^2 \right) \Delta \hat{\mathbf{U}}_{i,B}^l \\
& = \sum_{i=1}^{(N+1)^2} (\bar{L}_{il}^B)^2 \alpha_l^2 = \left\| \bar{\mathbf{L}}_l^B \alpha_l \right\|_2^2. \quad (12)
\end{aligned}$$

The same discretization is also applied for the second multi-dimensional integral in equation (11). For simplicity, we introduce PRSS in the following matrix notation:

$$PRSS \approx \left\| \dot{\mathbf{X}} - \bar{\mathbf{A}}\boldsymbol{\theta} \right\|_2^2 + \sum_{l=1}^{d^B} \lambda_l \left\| \bar{\mathbf{L}}_l^B \alpha_l \right\|_2^2 + \sum_{m=1}^{d^C} \mu_m \left\| \bar{\mathbf{L}}_m^C \beta_m \right\|_2^2, \quad (13)$$

where $\dot{\mathbf{X}} = (\dot{X}_1, \dot{X}_2, \dots, \dot{X}_N)^T$, $\bar{\mathbf{A}} = (\bar{\mathbf{A}}_1^T, \bar{\mathbf{A}}_2^T, \dots, \bar{\mathbf{A}}_N^T)^T$, $\boldsymbol{\theta} = (\boldsymbol{\alpha}^T, \boldsymbol{\beta}^T)^T$, $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \alpha_2, \dots, \alpha_{d^B})^T$ and $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \dots, \beta_{d^C})^T$, $\bar{G}_i + \bar{F}_i c_i = \bar{\mathbf{A}}_i \boldsymbol{\theta}$, $\bar{\mathbf{L}}_l^B = (L_{1l}^B, L_{2l}^B, \dots, L_{(N+1)2l}^B)^T$ and $\bar{\mathbf{L}}_m^C = (L_{1m}^C, L_{2m}^C, \dots, L_{(N+1)2m}^C)^T$ for $l = 1, 2, \dots, d^B$ and for $m = 1, 2, \dots, d^C$, respectively [21].

Using uniform penalization by taking the same λ for each derivative term, the regularized approximation problem of PRSS turns into a *Tikhonov regularization (TR)* problem:

$$PRSS \approx \left\| \dot{\mathbf{X}} - \bar{\mathbf{A}}\boldsymbol{\theta} \right\|_2^2 + \lambda \left\| \bar{\mathbf{L}}\boldsymbol{\theta} \right\|_2^2. \quad (14)$$

Here, $\lambda = \lambda_1 = \dots = \lambda_{d^B} = \mu_1 = \dots = \mu_{d^C}$, and $\bar{\mathbf{L}}$ is an $(M_{max} + 1) \times (M_{max} + 1)$ -diagonal matrix with first column $\mathbf{L}_0 = \mathbf{0}_{(N+1)^2}$ and the other columns being the vectors \mathbf{L}_l^B , \mathbf{L}_m^C , introduced above, where $M_{max} = d^B + d^C$.

As we just mentioned in Subsection 3.1, TR problem can be solved by a CQP program as given in equation (7). In order to write the optimality condition for this problem, we firstly reformulate our program as the subsequent *primal problem*:

$$\begin{aligned} & \min_{t, \boldsymbol{\theta}} \quad t, \\ & \text{such that} \quad \boldsymbol{\chi} := \begin{bmatrix} \mathbf{0}_N & \bar{\mathbf{A}} \\ 1 & \mathbf{0}_{M_{max}+1}^T \end{bmatrix} \begin{bmatrix} t \\ \boldsymbol{\theta} \end{bmatrix} + \begin{bmatrix} -\dot{\mathbf{X}} \\ 0 \end{bmatrix}, \\ & \boldsymbol{\eta} := \begin{bmatrix} \mathbf{0}_{M_{max}+1} & \bar{\mathbf{L}} \\ 0 & \mathbf{0}_{M_{max}+1}^T \end{bmatrix} \begin{bmatrix} t \\ \boldsymbol{\theta} \end{bmatrix} + \begin{bmatrix} \mathbf{0}_{M_{max}+1} \\ \sqrt{\bar{M}} \end{bmatrix}, \\ & \boldsymbol{\chi} \in L^{N+1}, \boldsymbol{\eta} \in L^{M_{max}+2}. \end{aligned} \quad (15)$$

Here, L^{N+1} , $L^{M_{max}+2}$ are the $(N+1)$ - and $(M_{max}+2)$ -dimensional *ice-cream* (or *second-order*) *cones* [15]. The *dual problem* to the latter problem is given

by

$$\begin{aligned}
& \max \quad (\dot{\mathbf{X}}^T, 0) \boldsymbol{\omega}_1 + \left(\mathbf{0}_{M_{max}+1}^T, -\sqrt{M} \right) \boldsymbol{\omega}_2, \\
& \begin{bmatrix} \mathbf{0}_N^T & 1 \\ \bar{\mathbf{A}}^T & \mathbf{0}_{M_{max}+1}^T \end{bmatrix} \boldsymbol{\omega}_1 + \begin{bmatrix} \mathbf{0}_{M_{max}+1}^T & 0 \\ \bar{\mathbf{L}}^T & \mathbf{0}_{M_{max}+1} \end{bmatrix} \boldsymbol{\omega}_2 = \begin{bmatrix} 1 \\ \mathbf{0}_{M_{max}+1} \end{bmatrix}, \\
& \boldsymbol{\omega}_1 \in L^{N+1}, \quad \boldsymbol{\omega}_2 \in L^{M_{max}+2}. \tag{16}
\end{aligned}$$

A primal-dual optimal solution $(t, \boldsymbol{\theta}, \boldsymbol{\chi}, \boldsymbol{\eta}, \boldsymbol{\omega}_1, \boldsymbol{\omega}_2)$ is obtained when the optimality conditions given in equation (17) are satisfied:

$$\begin{aligned}
\boldsymbol{\chi} &:= \begin{bmatrix} \mathbf{0}_N & \bar{\mathbf{A}} \\ 1 & \mathbf{0}_{M_{max}+1}^T \end{bmatrix} \begin{bmatrix} t \\ \boldsymbol{\theta} \end{bmatrix} + \begin{bmatrix} -\dot{\mathbf{X}} \\ 0 \end{bmatrix}, \\
\boldsymbol{\eta} &:= \begin{bmatrix} \mathbf{0}_{M_{max}+1} & \bar{\mathbf{L}} \\ 0 & \mathbf{0}_{M_{max}+1}^T \end{bmatrix} \begin{bmatrix} t \\ \boldsymbol{\theta} \end{bmatrix} + \begin{bmatrix} \mathbf{0}_{M_{max}+1} \\ \sqrt{M} \end{bmatrix}, \\
& \begin{bmatrix} \mathbf{0}_N^T & 1 \\ \bar{\mathbf{A}}^T & \mathbf{0}_{M_{max}+1}^T \end{bmatrix} \boldsymbol{\omega}_1 + \begin{bmatrix} \mathbf{0}_{M_{max}+1}^T & 0 \\ \bar{\mathbf{L}}^T & \mathbf{0}_{M_{max}+1} \end{bmatrix} \boldsymbol{\omega}_2 = \begin{bmatrix} 1 \\ \mathbf{0}_{M_{max}+1} \end{bmatrix}, \\
& \boldsymbol{\omega}_1^T \boldsymbol{\chi} = 0, \quad \boldsymbol{\omega}_2^T \boldsymbol{\eta} = 0, \\
& \boldsymbol{\omega}_1 \in L^{N+1}, \quad \boldsymbol{\omega}_2 \in L^{M_{max}+2}, \quad \boldsymbol{\chi} \in L^{N+1}, \quad \boldsymbol{\eta} \in L^{M_{max}+2}. \tag{17}
\end{aligned}$$

4. Application and Results

In order to test the theory developed in the previous section, we start with simulating stochastic process for a fixed Hurst parameter H using Cholesky method [4]. Now, our aim is to estimate the exact value of this Hurst parameter of the simulated data. We generate various stochastic processes which are the strong solution of SDEs driven by fBm with different Hurst parameters. Next, we construct CMARS model for each generated process to find the best fit. For the implementation of CMARS algorithm, BF's are built using Salford MARS[®] software program [11] as in [12, 22]. The optimization problem

given in equation (15) is solved by using interior point methods (IPMs) via the optimization software MOSEK [14, 16]. Finally, we examine the performances of CMARS fits according to well-known performance measures such as mean absolute error (MAE), mean squared error (MSE), correlation coefficient (r), multiple coefficient of determination (R^2), adjusted R^2 (Adj- R^2), and proportion of residuals within three sigma (PWI). The steps described above are applied for $H=0.2$, $H=0.3$, $H=0.7$, and $H=0.8$. The results of the applications are summarized in Table 1.

Table 1: CMARS performances for fBm generated by $H=0.2$, $H=0.3$, $H=0.7$, and $H=0.8$.

Hurst index	Performance Measures					
	MAE	MSE	r	R^2	Adj- R^2	PWI
$H = 0.1$	0,8766	1,4827	0,1923	0,0370	-0,1651	1
$H = 0.2$	0,6207*	0,7480*	0,9868*	0,9739*	0,9684*	1
$H = 0.3$	0,8861	1,5266	0,0991	0,0098	-0,1979	1
$H = 0.4$	0,8770	1,4733	0,2075	0,0430	-0,1577	1
$H = 0.5$	0,8839	1,5201	0,1281	0,0164	-0,1900	1
$H = 0.1$	0,7138	0,9776	0,2607	0,0679	-0,1276	1
$H = 0.2$	0,7162	0,9901	0,2400	0,0576	-0,1401	1
$H = 0.3$	0,3606*	0,2516*	0,984*	0,9699*	0,9636*	1
$H = 0.4$	0,6926	0,9387	0,3249	0,1055	-0,0821	1
$H = 0.5$	0,7053	0,9763	0,2641	0,0697	-0,1254	1
$H = 0.5$	0,7031	0,9719	0,4743	0,2250	0,0623	1
$H = 0.6$	0,7048	0,9784	0,4688	0,2198	0,0560	0,9898
$H = 0.7$	0,3634*	0,2602*	0,9582*	0,9182*	0,9010*	1
$H = 0.8$	0,7041	0,9506	0,4919	0,2419	0,0828	1
$H = 0.9$	0,7081	0,9781	0,4691	0,2200	0,0563	1
$H = 0.5$	0,6068	0,7841	0,5914	0,3498	0,2133	1
$H = 0.6$	0,6359	0,8015	0,5783	0,3345	0,1948	1
$H = 0.7$	0,6053	0,7389	0,6176	0,3815	0,2517	1
$H = 0.8$	0,2009*	0,0822*	0,9883*	0,9768*	0,9720*	1
$H = 0.9$	0,6006	0,7294	0,6240	0,3894	0,2613	1

* indicates better performance

In the case for anti-persistent processes, namely, $H = 0.2$, $H = 0.3$, the values of MAE and MSE are lower and the values of R^2 , Adj- R^2 , PWI and r are higher than the values for the other Hurst parameter values. Similar results are also obtained for the case of persistent processes. Hence, this shows that according to performance measures criteria, the best CMARS fit gives us the correct Hurst parameter value.

5. Conclusion and Outlook

Recent developments in computer science provide environments in order to collect numerous data from various sources. Data mining methods enable us to analyze data for different purposes in many fields, such as finance, environment, and energy. One of the modern method of data mining, CMARS, has been developed as an alternative to the backward stepwise part of the MARS algorithm (see [22]).

This paper gave a new contribution to Hurst parameter estimation theory for the strong solution of SDEs driven by fBm using CMARS technique. The main superiority of our approach to the others is that it not only estimates the Hurst parameter but it also finds spline parameters of the stochastic process. What is more, our representation of financial and other processes is empowered by all the modeling and numerical advantages of CMARS. By this, a bridge has been offered between convex optimization and Hurst parameter estimation theory.

In this pioneering paper, we followed a *two-level* approach with the determination of the parameters at the *lower level*, except of the Hurst-parameter which was chosen at the following *upper level*. This approach can be regarded as a *parametric optimization* (cf. [7, 9]). In future research, we will deepen and extend this approach by both more *model-free* strategies (e.g., from statistics and data mining), especially, more *model-based* ones, and with a comparison of them. The model-based approaches will be of a more *integrated* mathematical nature and in the analytical line that we initiated in this work. Through these investigations we intend to further contribute to a deeper understanding of our modern financial markets and to offer helpful mathematical decision tools for them.

Acknowledgement

Fatma Yerlikaya-Özkurt is supported by the TUBITAK Domestic Doctoral Scholarship Program.

References

- [1] B.-E. Alos, O. Mazet, D. Nualart, Stochastic calculus with respect to Gaussian processes, The Annals of Probability, 29(2) (2001) 766-801.
- [2] F. Biagini, Y. Hu, B. Øksendal, Stochastic Calculus for Fractional Brownian Motion and Applications, Springer-Verlag, London, 2008.

- [3] A. Chronopoulou, F.-G. Viens, Hurst index estimation for self-similar processes with long-memory, in: J. Duan, S. Luo, C. Wang (Eds.), Recent Advances in Stochastic Dynamics and Stochastic Analysis, World Scientific Publishing Co Pte Ltd, 2009, pp. 1-28.
- [4] T. Dieker, Simulation of fractional Brownian motion, MSc. Thesis, University of Twente, Amsterdam, 2004.
- [5] J.-H. Friedman, Multivariate adaptive regression splines, The Annals of Statistics, 19, (1991) 1-141.
- [6] M. Gradinaru, I. Nourdin, Milstein's type schemes for fractional SDEs, Annales de l' Institut Henri Poincaré-Probabilités et Statistiques, 45(4) (2009) 1085-1098.
- [7] J. Guddat, F. Guerra, H.-Th. Jongen, Parametric Optimization: Singularities, Pathfollowing and Jumps, John Wiley & Sons, BG Teubner, Stuttgart, Chichester, 1990.
- [8] H. Hurst, Long term storage capacity of reservoirs, Transactions of the American Society of Civil Engineers, 116 (1951) 770-799.
- [9] H.-Th. Jongen, G.-W. Weber, On parametric nonlinear programming, Annals of Operations Research 27 (1990) 253-284.
- [10] P.-E. Kloeden, E. Platen, H. Schurz, Numerical Solution of SDE Through Computer Experiments, Springer Verlag, New York, 1994.
- [11] MARS from Salford Systems. Available at <http://www.salfordsystems.com/mars/phb>.
- [12] MATLAB Version 7.5 (R2007b)
- [13] Y. Mishura, Stochastic Calculus for Fractional Brownian Motion and Related Topics, Springer-Verlag, Berlin, Heidelberg, 2008.
- [14] MOSEK, A very powerful commercial software for CQP. Available at <http://www.mosek.com>.
- [15] A. Nemirovski, A lectures on modern convex optimisation, Israel Institute of Technology, 2002.
<http://iew3.technion.ac.il/Labs/Opt/opt/LN/Final.pdf>.

- [16] Y.-E Nesterov, A.-S Nemirovski, Interior Point Polynomial Algorithms in Convex Programming, SIAM, 1994.
- [17] B.-L.-S. Rao, Statistical Inference for Fractional Diffusion Processes, first ed., UK, 2010.
- [18] A. Sharkasi, M. Crane, H.-J. Ruskin, J.-A. Matos, The Reaction of Stock Markets to Crashes and Events: A Comparison Study between Emerging and Mature Markets using Wavelet Transforms, *Physica A*, 368 (2006) 511-521.
- [19] P. Taylan, G.-W. Weber, Organization in finance prepared by stochastic differential equations with additive and nonlinear models and continuous optimization, *Organizacija (Organization - Journal of Management, Information Systems and Human Resources)*, 41(5) (2008) 185-193.
- [20] F. Yerlikaya, A new contribution to nonlinear robust regression and classification with MARS and its application to data mining for quality control in manufacturing, MSc. Thesis, Middle East Technical University, Ankara, 2008.
- [21] F. Yerlikaya-Özkurt, G.-W. Weber, Parameter estimation of stochastic differential equations with conic multivariate adaptive regression splines method, working paper, IAM, METU, 2013.
- [22] G.-W. Weber, İ. Batmaz, G. Köksal, P. Taylan, F. Yerlikaya-Özkurt, CMARS: A new contribution to nonparametric regression with multivariate adaptive regression splines supported by continuous optimisation, *Inverse Problems in Science and Engineering*, 20 (2011) 371-400.